

## A collective intelligence based contents search method for massive contents

Ito, Eisuke  
Research Institute for Information Technology

Hirokawa, Sachio  
Research Institute for Information Technology

<https://hdl.handle.net/2324/25117>

---

出版情報 : 2012-09-10. The 16th Core University Program Seminar on Next Generation Internet  
バージョン :  
権利関係 :

# A collective intelligence based contents search method for massive contents

Eisuke Ito and Sachio Hirokawa

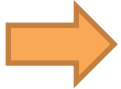
Kyushu University

[ito.eisuke.523@m.kyushu-u.ac.jp](mailto:ito.eisuke.523@m.kyushu-u.ac.jp)

[hirokawa@cc.kyushu-u.ac.jp](mailto:hirokawa@cc.kyushu-u.ac.jp)

# Outline

---



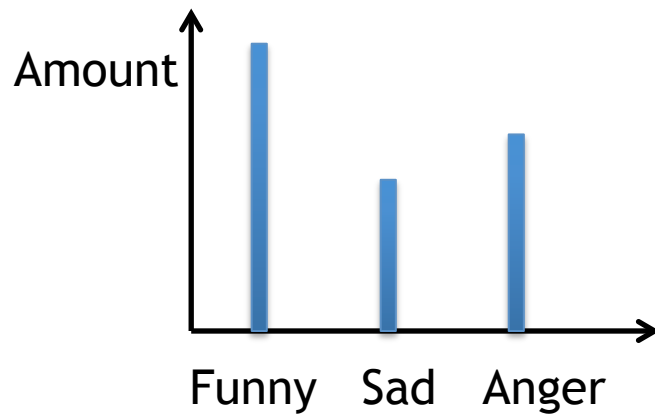
1. Introduction
2. Modeling
3. Nicovideo
4. Ranking by comments
5. Tag Hierarchy
6. Conclusion

# 1. Introduction

- Massive contents are created
  - Professionals, Semi-pro and The commons
- Type of contents
  - Movie, Music: YouTube, Nicovideo, Youku
  - Photo, Picture, 3D object
  - Cartoon, Comic, Manga
  - Novel
  - News, Blog article, ...
  - Academic article: Institutional Repository
- Q: **How to find good content?**
- A: Use collective intelligence
  - Folksonomy, Social intelligence, ...
    - Users comments, tags,



## Emotion / Sympathy



How to quantify?

## Category / Clustering

Music	Sports
Classic	Ball games
Rock	Football, Baseball, ...
Jazz	Martial arts
Sound track	Judo, TaeKwonDo...
	Field and tracks

How to find category?  
How to make hierarchy?

The screenshot shows the YouTube homepage interface. At the top left is the YouTube logo. To its right is a search bar with a 'Search' button. Further right are links for 'Browse' and 'Upload'. Below the search bar is a section titled 'All Categories' with a dropdown menu currently set to 'Categories'. Below this, there is a grid of category links:

All	Entertainment	Music	Pets & Animals
Autos & Vehicles	Film & Animation	News & Politics	Science & Technology
Comedy	Gaming	Nonprofits & Activism	Sports
Education	Howto & Style	People & Blogs	Travel & Events

# Objectives

---

Ranking by Viewers emotion  
(Ranking by comments)

Automatic categorizing  
(Tag Hierarchy)

# Outline

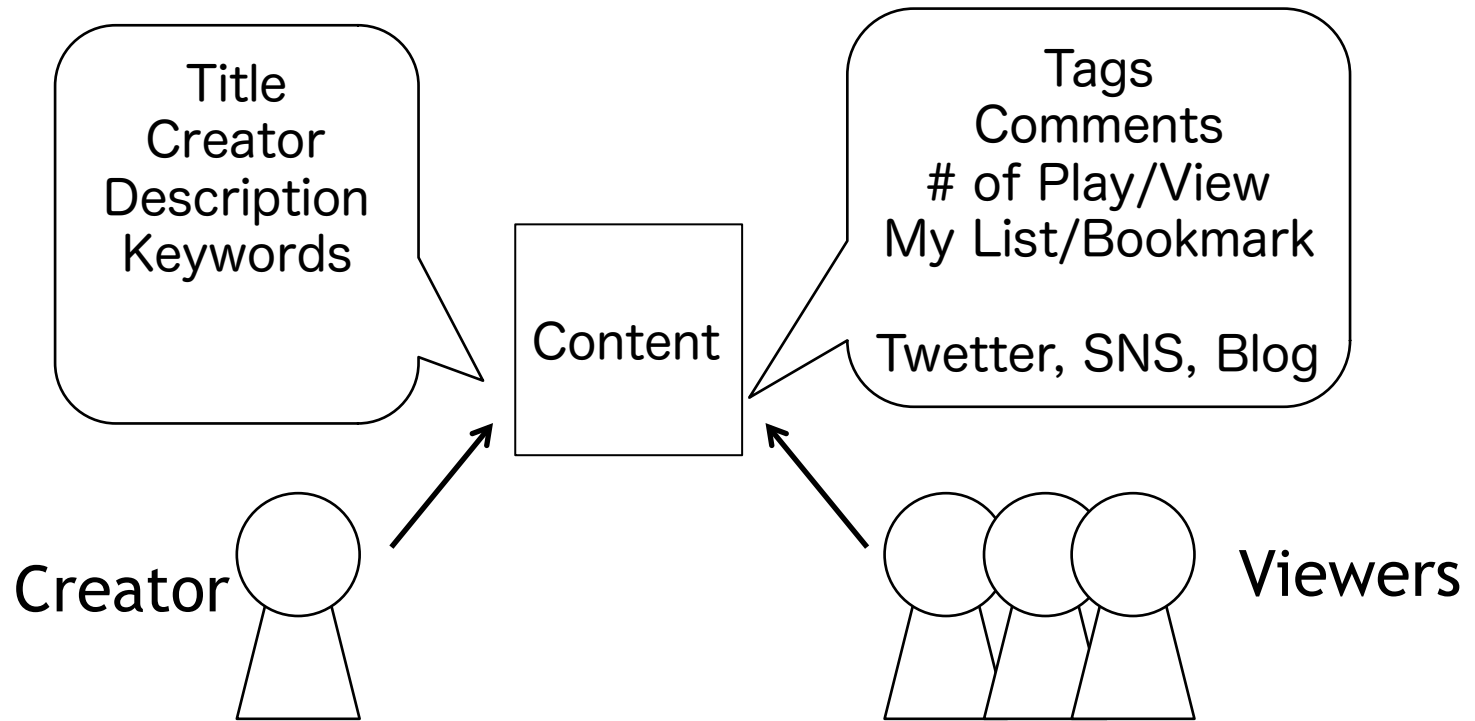
---



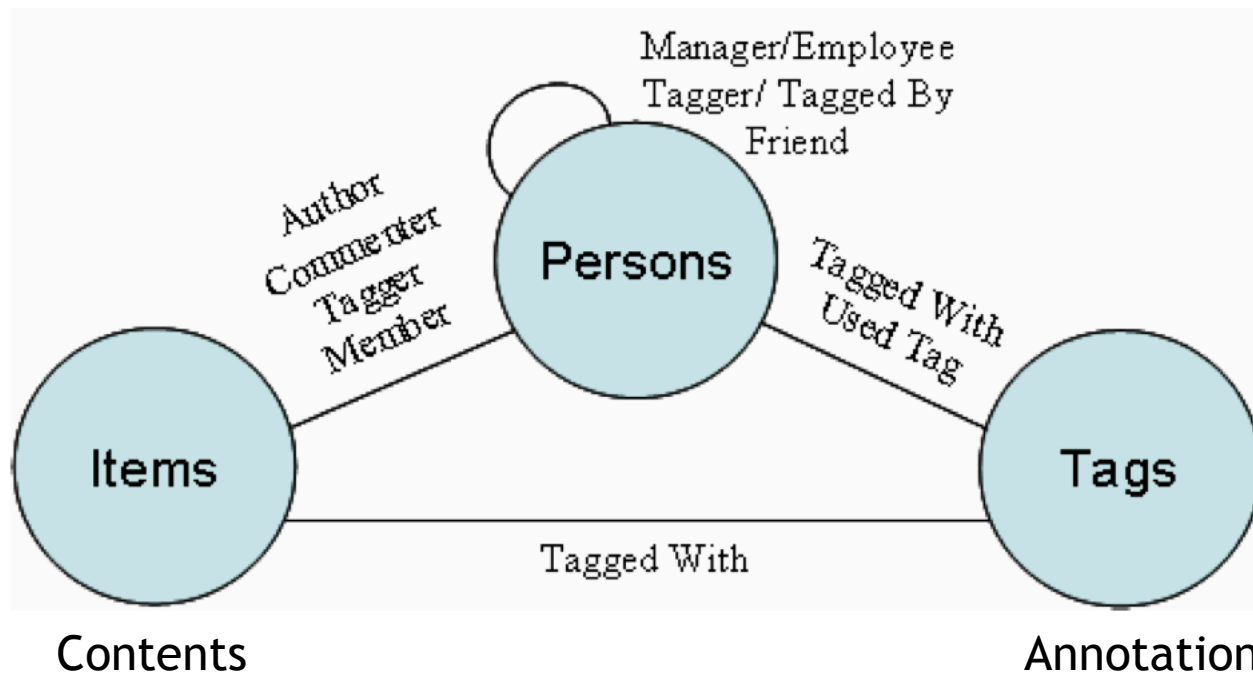
1. Introduction
2. Modeling
3. Nicovideo
4. Ranking by comments
5. Tag Hierarchy
6. Conclusion



## 2. Modeling

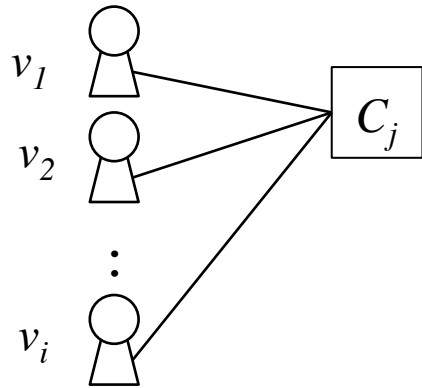


# Three objects

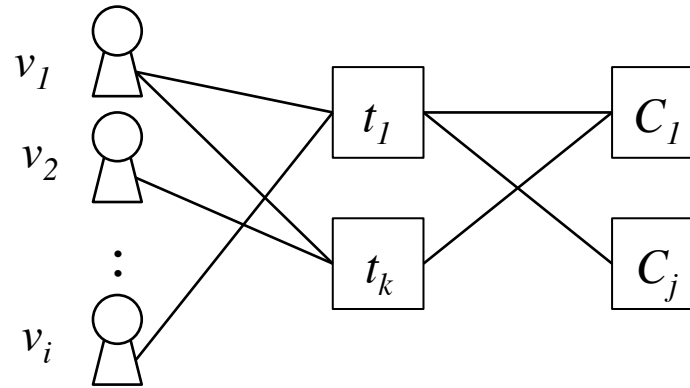


I. Guy, N. Zwerdling, I. Ronen, D. Carmel, E. Uziel:  
Social media recommendation based on people and tags,  
ACM SIGIR2010, pp.194-201, 2010.

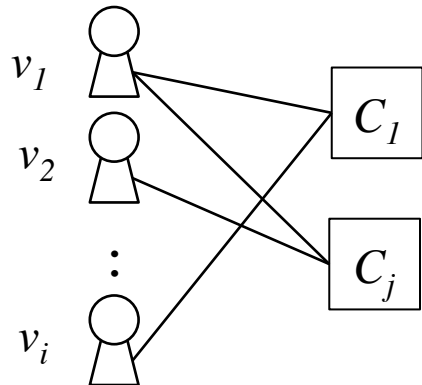
Viewers Content



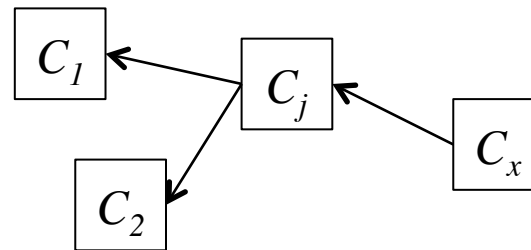
Viewers Tags Contents



Viewers Contents



Contents Links (Reference)



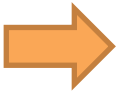
Creators network

Viewers community

# Outline

---

1. Introduction
2. Modeling
3. Nicovideo
4. Ranking by comments
5. Tag Hierarchy
6. Conclusion



# 3. Nicovideo

- Movie uploading/streaming service in Japan.
- Popular for Geeks (OTAKU).
  - Animation, Computer games, ...
- More than 5 million movies (Feb., 2011)
- # of Users

Class	# of ID	Description
Normal	17 million	Free, Low QoS
Premium	1 million	Charged, High QoS

# Unique comment system

- Viewer can submit his/her comments
- Comments flow on the movie according to the submit time.
- Illusionary effect
  - Watch with other viewers

The screenshot shows a video player interface. The video content displays two people in a car, with overlaid text comments such as "この躍動感すごい" and "みっくみっくに". The right side of the interface features a comment list with columns for "再生時" (Playback Time) and "コメント" (Comment). The list shows comments submitted at various times, such as "同期w" at 00:37 and "みっくみっくに" at 00:40. The interface also includes a progress bar at the bottom and a "再生" (Play) button.

再生時	コメント
00:37	同期w
00:37	みくかわyusu
00:37	スタンドwwwwww
00:37	見えない・・・
00:37	かあい
00:38	モデルじゃねえの？
00:38	テラカワユス
00:38	謎の技術wwwwww
00:39	すごええええええええ
00:39	この躍動感すごい
00:39	みっくみーくにしてあげる♪
00:39	ここまでミクとシンクロするとは
00:40	カチッととまるのね～
00:40	ちゃんと張りあえてる
00:40	みっくみっくに

# Outline

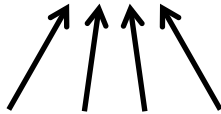
---

1. Introduction
2. Modeling
3. Nicovideo
4. Ranking by comments
5. Tag Hierarchy
6. Conclusion



# 4. Ranking by comments

$C_x$



Nice!  
I like it!  
Wao!

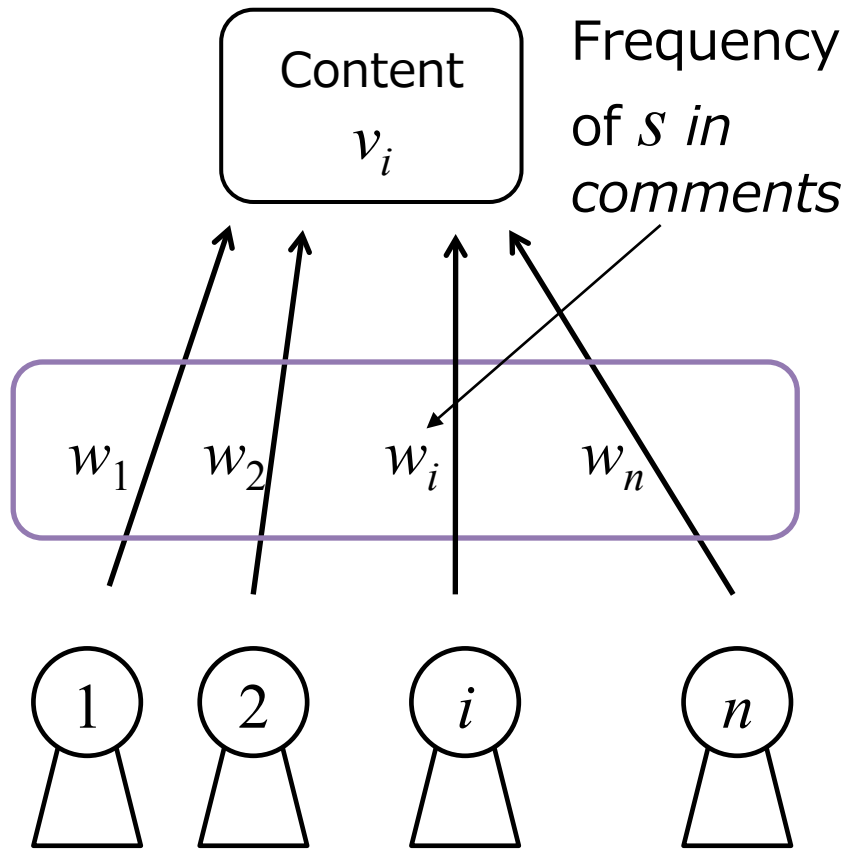
$C_y$



...



# Ranking by comments



$n$ : # of commented viewers

Weight of commenter  $i$

$$w_i = \log(\text{freq}_i(s) + 1)$$

$s$ : feature word(s)

$W(v)$ : weight of  $v$

$$W(v) = \frac{\sum w_i}{n}$$

# Implementation

**SeeMore**  
ニコニコ動画てきとう検索エンジン

平沢進

平沢進の検索結果751件中1..32件

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

Target Data: “Music” tagged Movies  
on Nicovideo

# of movies: 373,265



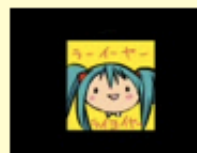
[\[ゆっくり\] LAB=01 \[P-MOD...](#)

0.661776310647043



[イタリアでいちばんピ...](#)

0.535791787329755



[平沢進と童歌](#)

0.440182593502529



[平沢進と童歌](#)

0.440182593502529



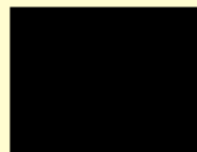
[susmoooch` d`](#)

0.423468844705594



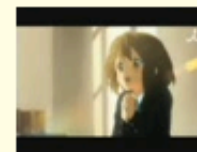
[\[歌唱代理人\]初音ミクで...](#)

0.415360642965562



[忙しい人のための「As...](#)

0.41007299582182



[けいおん！PVにP-MODELの...](#)

0.374984139194189

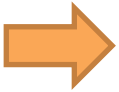
# Results

- We can find good(interesting) movies.
- But...
  - MAD movies
  - They are not general.
  - Contents for small community
- Use other data
  - replay number, bookmarked users, ...
- Try other models
  - Viewers-Contents bipartite graph
  - Apply HITS algorithm

# Outline

---

1. Introduction
2. Modeling
3. Nicovideo
4. Ranking by comments
5. Tag Hierarchy
6. Conclusion



# 4. Tag hierarchy

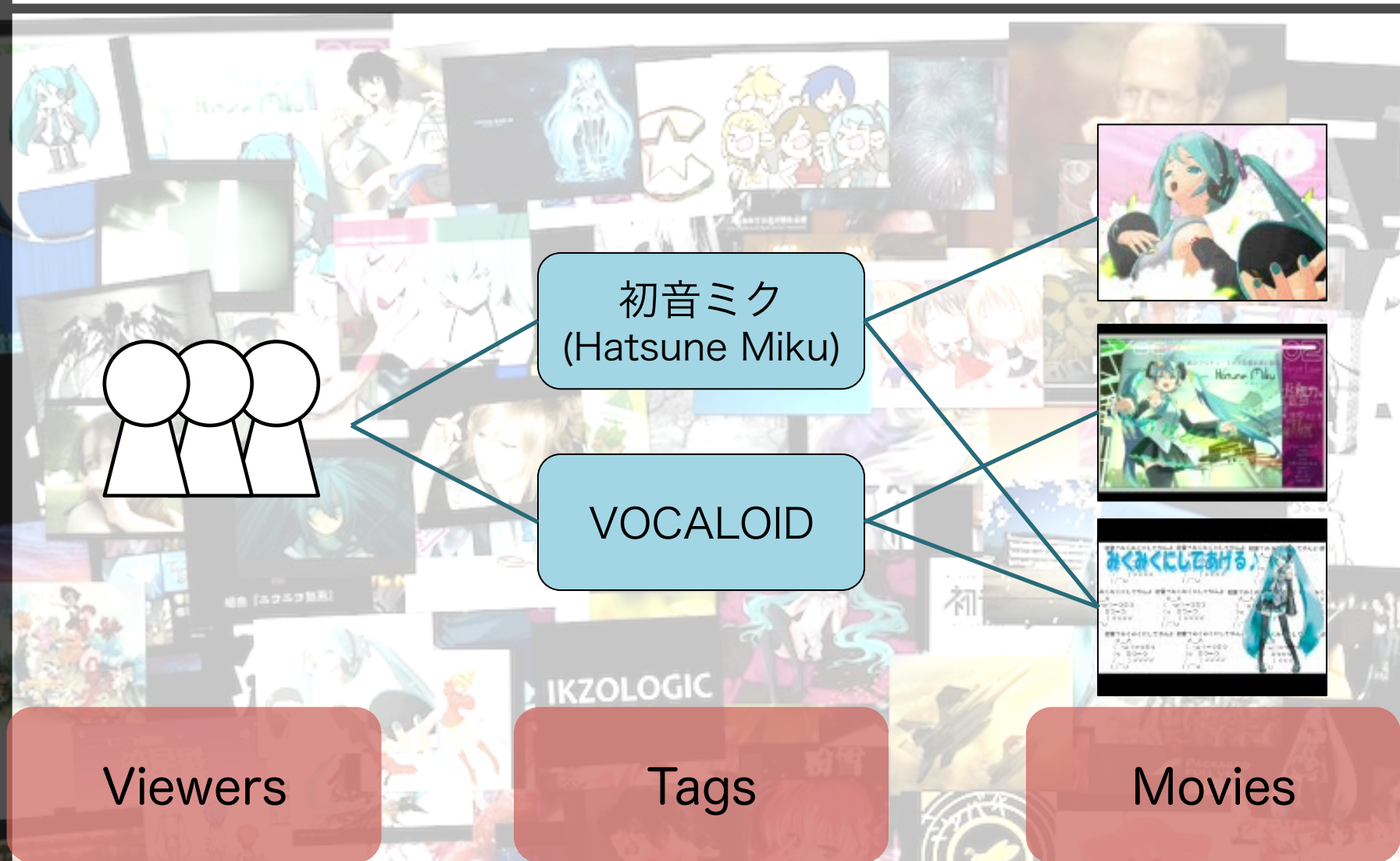
---

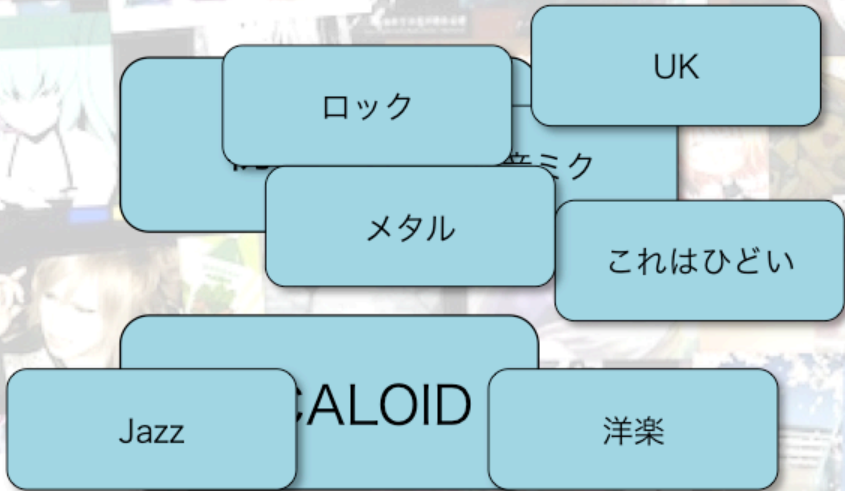
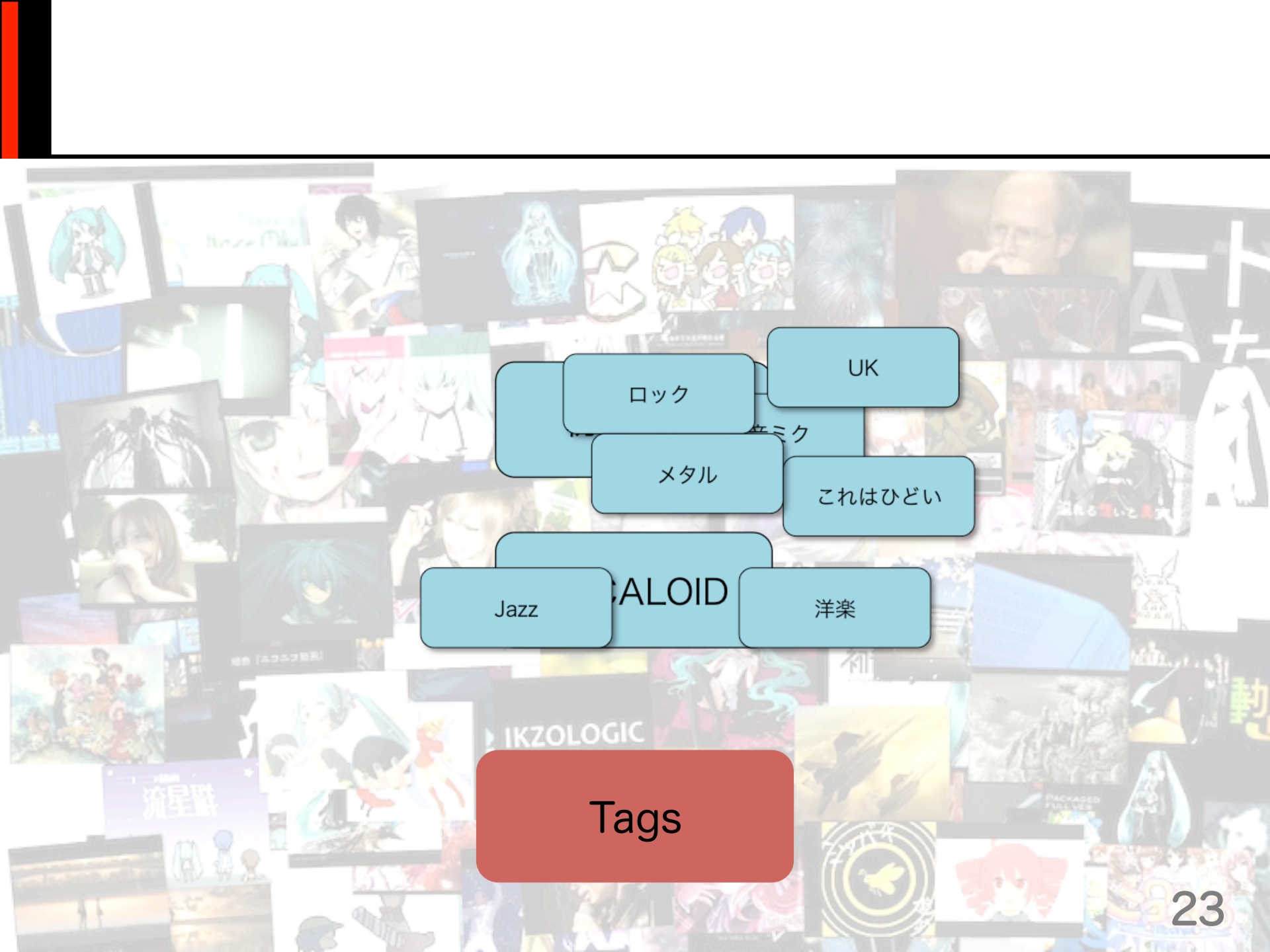
- Huge amount of movies
- Given categories are not enough
- Search by tag.
  - Tagging by viewers
  - Huge number of tags.
  - What is appropriate tag for my uncertain request?
- Automatic tag hierarchy generation

# Huge amount of movies



# Tagging by viewers





Tags



誰得

映画音楽

イギーポップ

ユニコーン

The\_wh

誰得

ゲーム

The\_pogues

UK

インド

FF5

Floyd

パンク

鏡音レン

作業妨害用BGM

QUEEN

The\_beatles

迷曲

初音

トランス

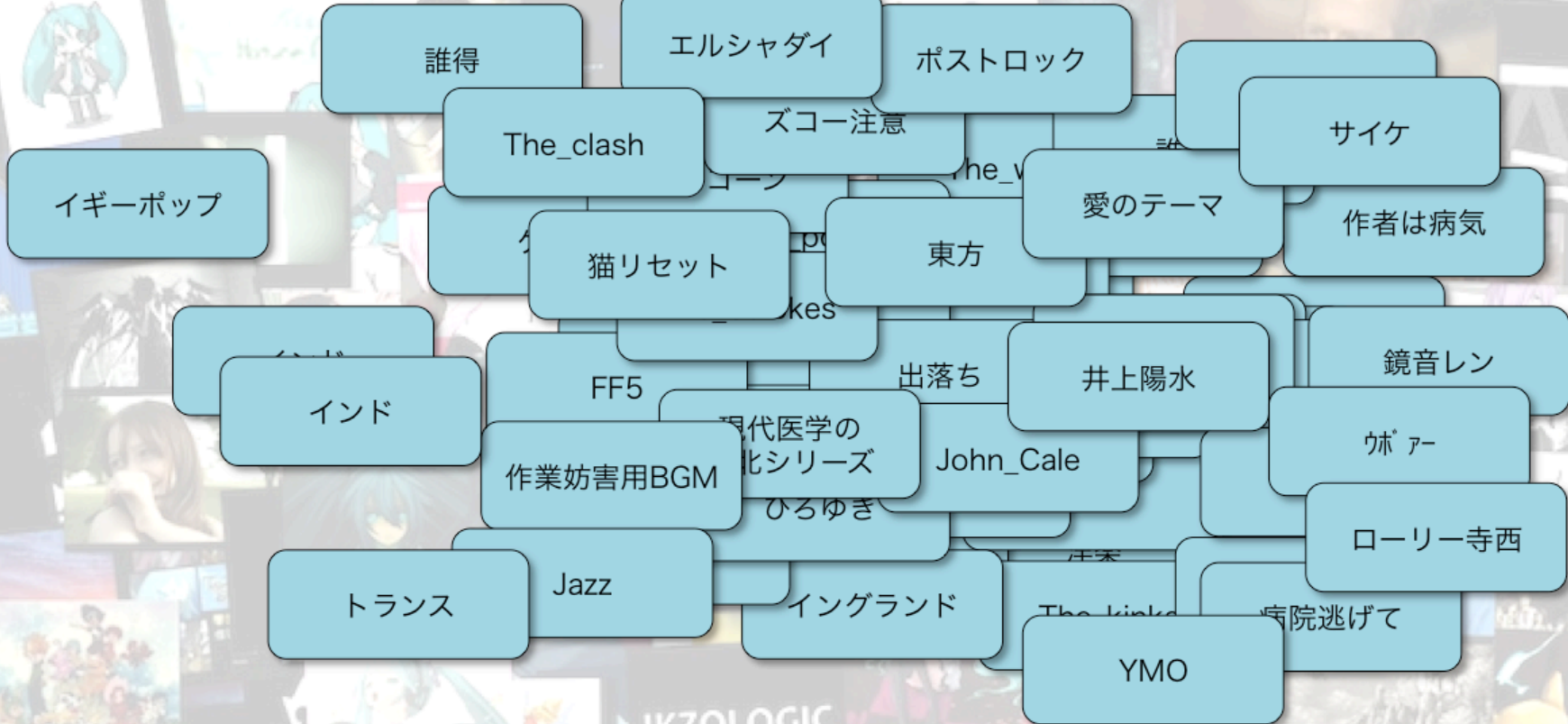
Jazz

イングランド

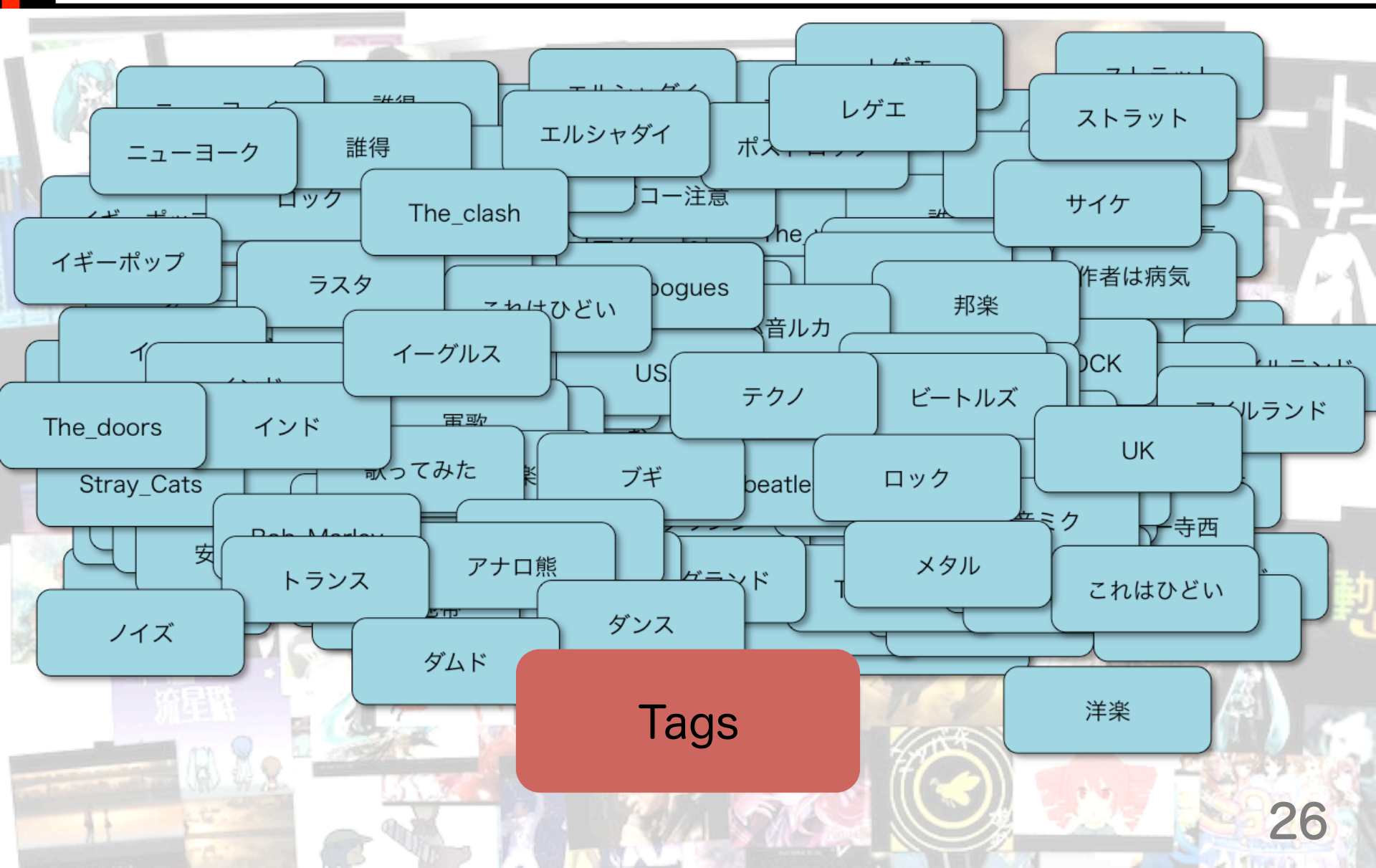
The\_kinks

ELO

# Tags



Tags



ニューヨーク

誰得

エルシャダイ

レゲエ

ストラット

イギーポップ

The\_clash

サイケ

ラスタ

bogues

邦楽

作者は病気

イーグルス

音ルカ

The\_doors

インド

US

テクノ

ビートルズ

DCK

アイルランド

Stray\_Cats

歌ってみた

ブギ

beatle

ロック

UK

安

Bob\_Marley

トランス

アナロ熊

ガレンド

メタル

音ミク

寺西

ノイズ

ダンス

ダムド

ダンス

これはひどい

Tags

洋楽

What is appropriate tag  
for my uncertain request?

タグ

洋楽

これはひどい

メタル

ダンス

ダムド

ノイズ

トランス

アナロ熊

ブギ

Stray\_Cats

The\_doors

UK

ロック

beatle

歌ってみた

Bob\_Marley

安

辛mik

寺西

ガレンド

アイルランド

US

イーグルス

ラスト

イギーポップ

The\_clash

コー注意

ロック

誰得

ニューヨーク

エルシャダイ

ポス

レゲエ

ストラット

サイケ

作者は病気

bogues

地楽

イ

インド

雷駟

US

メトロリス

地楽

ロック

アイルランド

アイルランド

アイルランド

アイルランド

アイルランド

アイルランド

アイルランド

アイルランド

アイルランド

アイルランド

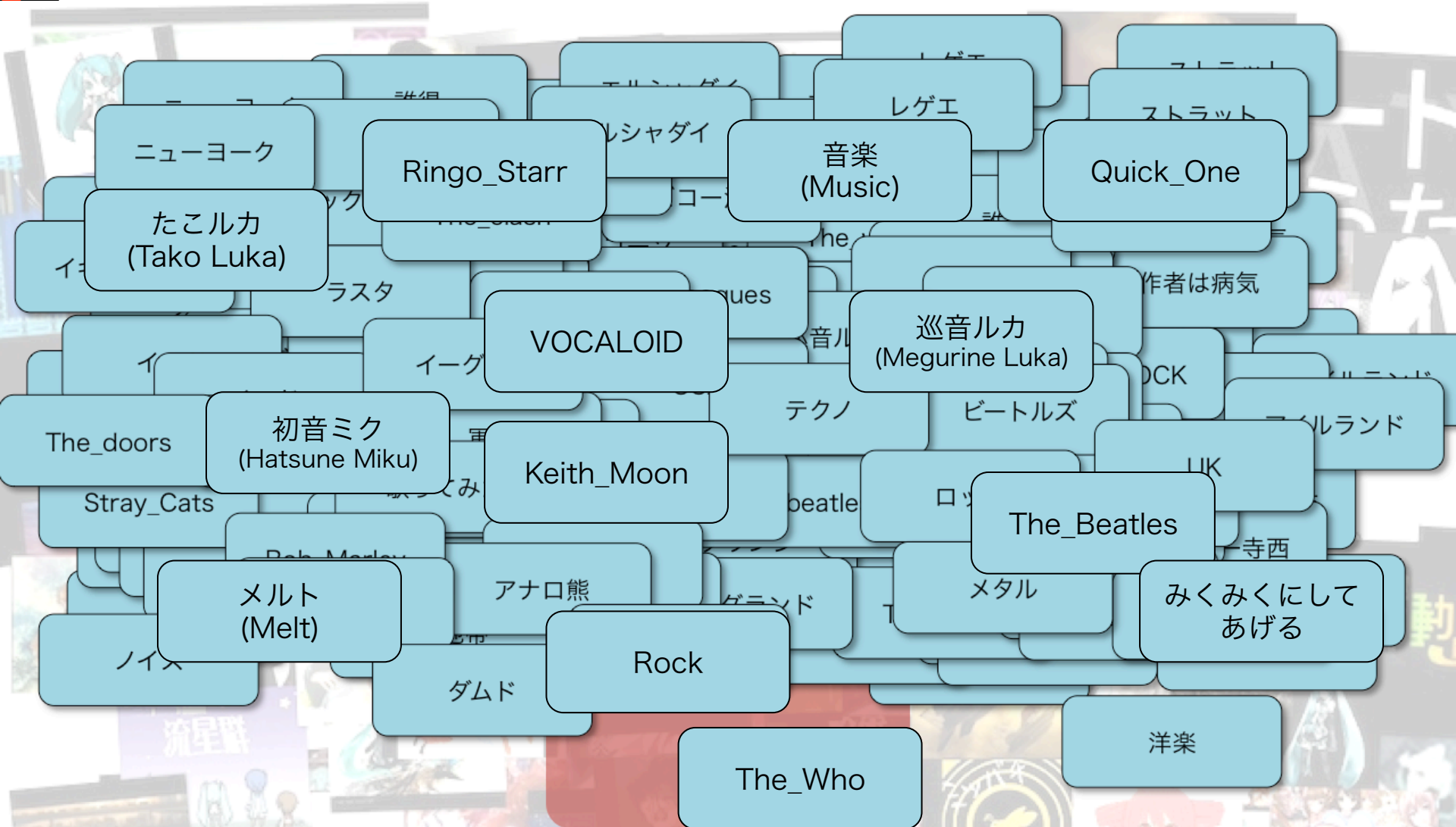
アイルランド

アイルランド

アイルランド

アイルランド

# Messy



ニューヨーク

Ringo\_Starr

音楽  
(Music)

Quick\_One

たこルカ  
(Tako Luka)

VOCALOID

巡音ルカ  
(Megurine Luka)

初音ミク  
(Hatsune Miku)

Keith\_Moon

The\_Beatles

メルト  
(Melt)

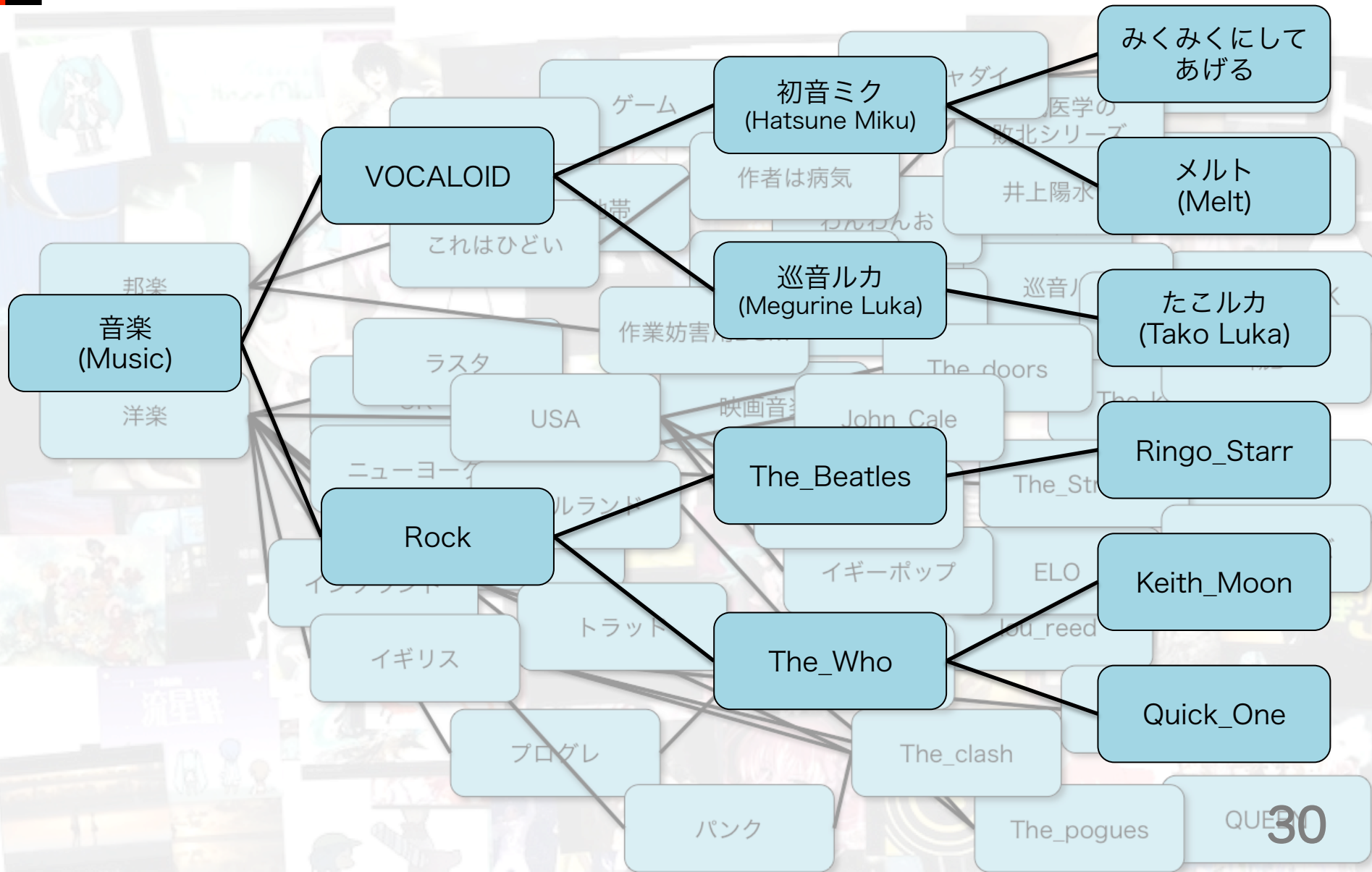
Rock

みくみくにして  
あげる

The\_Who

洋楽

# Clear!

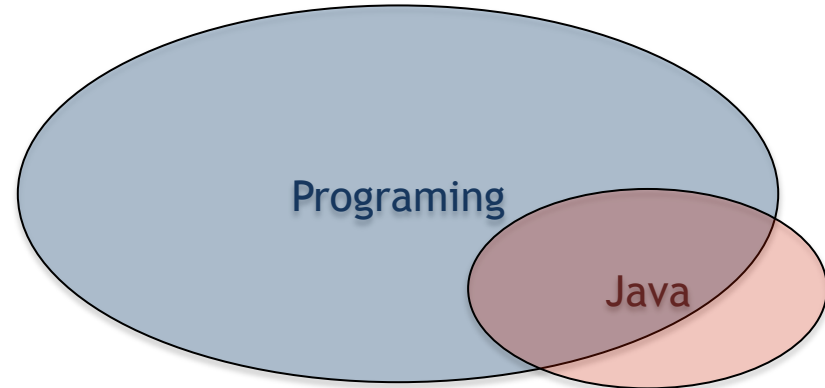
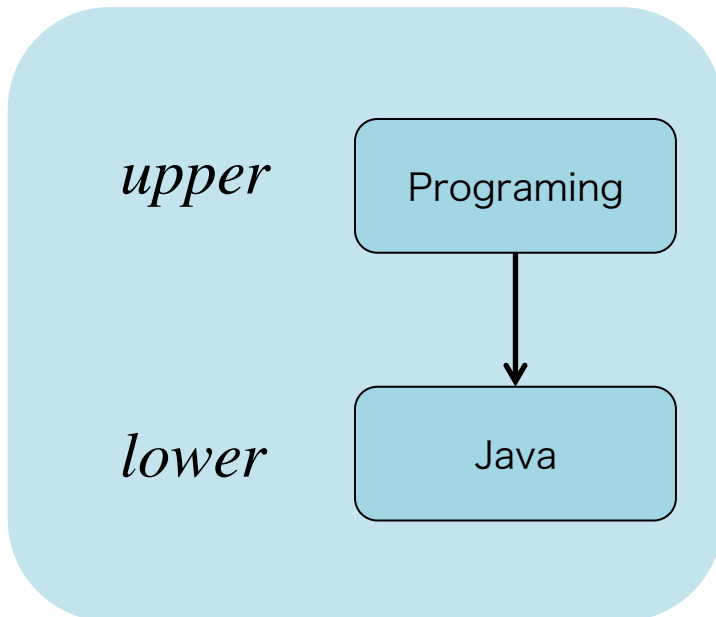


# Def. of upper term

*u is upper term of v*

Intersection Ratio

$$\Leftrightarrow D(u) > D(v) \quad \text{and} \quad \frac{|D(u) \cap D(v)|}{|D(v)|} > \alpha$$





# Implementation

**TagMore**

おもしろタグ発見ツール

Target Data: “Music” tagged Movies  
on Nicovideo  
# of movies: 373,265

Input a word,  
Shows upper words (tags) and lower words(tags)

### upper tags

The Beatles ビートルズ : 0.555

Western Music 洋楽 : 0.348

Niconico movie maker ニコニコムービーメーカー : 0.148

UK : 0.088

60's : 0.051

Live : 0.044

High resolution 高画質 : 0.037

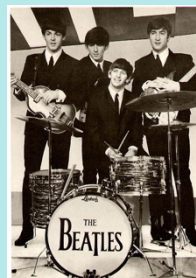
More Valid もっと評価されるべき : 0.029

BGM for work 作業用BGM : 0.029

Rock ロック : 0.029

### input

## beatles



### lower tags

ポール・マッカートニー : 0.088 Paul McCartney

ジョン・レノン : 0.051 John Lennon

Paul McCartney : 0.044

リンゴ・スター : 0.029 Ringo Starr

リマスター : 0.029 Re-master

John Lennon : 0.022

the : 0.022

THE BEATLES : 0.022

The Beatles : 0.022

Paul : 0.014

### upper tags

Western Music 洋楽 : 0.368

Comments by author 投稿者コメント : 0.146

BGM for work 作業用BGM : 0.076

Rock ロック : 0.07

BGM for work アニメ色のない作業用BGM : 0.052  
without animation taste

80's : 0.052

神曲 : 0.046

Niconico movie maker ニコニコムービーメーカー : 0.04

UK : 0.04

Recommend headphone ヘッドホン推奨 : 0.035

### input

## QUEEN



### lower tags

クイーン : 0.251 Queen

フレディ・マーキュリー : 0.093 Freddie Mercury

Bohemian Rhapsody : 0.081

ブライアン・メイ : 0.081 Brian May

Brian May : 0.07

QUEEN厳選動画 : 0.064

Queen名曲選 : 0.052

60fps : 0.052

フレディ : 0.052 Freddie

[LIVE] : 0.052

### upper tags

Niconico movie maker

ニコニコムービーメーカー : 0.218

Comments by author 投稿者コメント : 0.105

More valid もっと評価されるべき : 0.064

Recommend for @BGM @BGM推奨 : 0.042

High resolution 高音質 : 0.038

Mash-up マッシュアップ : 0.037

VOCALOID : 0.037

BGM for work 作業用BGM : 0.036

Hatsune Miku 初音ミク : 0.036

Vocaloid cover VOCALOIDカバー曲 : 0.026

### input

K-On!

けいおん!



### lower tags

Don't say "lazy" : 0.225

Cagayake!GIRLS : 0.143

ふわふわ時間 : 0.132 Fuwa-Fuwa time

平沢唯 : 0.117 Hirasawa Yui

秋山澪 : 0.105 Akiyama mio

桜高軽音部 : 0.072 Sakura-kou music club

キャラソン : 0.065 Chara. Son.

放課後ティータイム : 0.063 After school tea time

『レッツゴー』 : 0.059 "Let's go"

ほかおん! : 0.049 Voca-On!

### upper tags

けいおん! : 0.915

ふわふわ時間 : 0.294

平沢唯 : 0.263

秋山澪 : 0.178

投稿者コメント : 0.136

桜高軽音部 : 0.094

けいおん! : 0.084

ニコニコムービーメーカー : 0.073

作業用BGM : 0.063

高音質 : 0.052

### input

After school tea time

放課後ティータ  
イム



### lower tags

わたしの恋はホッチキス : 0.263

ふでペン ~ボールペン~ : 0.168

カレーのちライス : 0.105

田井中律 : 0.052

琴吹紬 : 0.052

かわ唯 : 0.052

『レッツゴー』 : 0.052

日笠陽子 : 0.042

豊崎愛生 : 0.042

中野梓 : 0.031



upper tags

ニコニコムービーメーカー : 0.084

Niconico movie maker

input

VOCALOID



lower tags

初音ミク : 0.539 Hatsune Miku

ミクオリジナル曲 : 0.286 Miku original

鏡音リン : 0.157 Kagamine Rin

VOCALOIDカバー曲 : 0.151 Vocaloid cover

鏡音レン : 0.122 Kagamine Len

KAITO : 0.119

巡音ルカ : 0.089 Megurine Luka

リンオリジナル曲 : 0.066 Rin original

もっと評価されるべき : 0.059 More valid

VOCAROCK : 0.054



upper tags

VOCALOID : 0.611

ニコニコムービーメーカー : 0.084

Niconico movie maker

作業用BGM : 0.018

BGM for work

input

(Hatsune Miku)

初音ミク



lower tags

ミクオリジナル曲 : 0.413 Miku original

VOCALOIDカバー曲 : 0.092 Vocaloid cover

鏡音リン : 0.067 Kagamine Rin

もっと評価されるべき : 0.06 More valid

VOCAROCK : 0.046

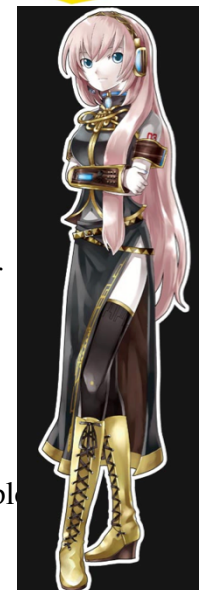
投稿者コメント : 0.043 Comments by author

おっさんホイホイ : 0.043 Old men capture

鏡音レン : 0.041 Kagamine Len

KAITO : 0.039

みんなのミクうた : 0.038 Miku's songs for people





Medvedev

upper tags

input

lower tags

Red Nicovideo 赤いニコニコ動画 : 0.6

Russia ロシア : 0.6

Classic クラシック : 0.2

National anthem 国歌 : 0.16

Soviet union ソ連 : 0.16

BGM for work 作業用BGM : 0.16

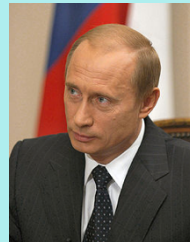
Anthems in the world 世界の国歌 : 0.16

War song 軍歌 : 0.16

祖国は我らのために : 0.12

Russian folk song ロシア民謡 : 0.08

Putin  
プーチン



メドヴェージェフ : 0.08

KGB専用便利リンク : 0.08

軍事パレード : 0.08

おそロシア : 0.08

グリンカ : 0.08

うp主は粛清されました : 0.04

改竄版 : 0.04

栄光あれ : 0.04

つくねちゃん : 0.04

聞けば聞くほどロシアが好きになる : 0.04

# Outline

---

1. Introduction
2. Modeling
3. Nicovideo
4. Ranking by comments
5. Tag Hierarchy
6. Conclusion



# 6. Conclusion

- Content search method using collective intelligence
  - Contents ranking based on viewers comments
  - Tag Hierarchy generation based on DF (document frequency) and intersection ratio
    - Not Tree, but DAG (direct acyclic graph)
- Ranking is not good.
  - Other information (# of replay, bookmarks)
  - Analyze relation between comments, tags, and viewers
  - Other graph models